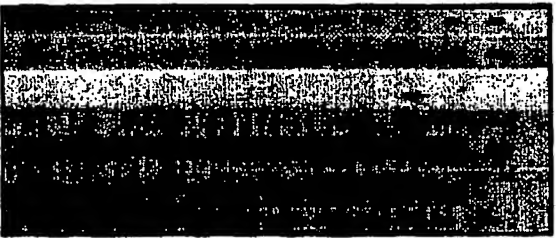


PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification <sup>6</sup> : <b>C12Q 1/68</b></p>	<b>A1</b>	<p>(11) International Publication Number: <b>WO 99/01576</b></p> <p>(43) International Publication Date: <b>14 January 1999 (14.01.99)</b></p>																																																																																				
<div style="display: flex; justify-content: space-between;"> <div style="width: 48%;"> <p>(21) International Application Number: <b>PCT/GB98/01940</b></p> <p>(22) International Filing Date: <b>2 July 1998 (02.07.98)</b></p> <p>(30) Priority Data:</p> <table style="width: 100%; border: none;"> <tr> <td style="width: 33%;">9714005.7</td> <td style="width: 33%;">2 July 1997 (02.07.97)</td> <td style="width: 33%;">GB</td> </tr> <tr> <td>9806702.8</td> <td>27 March 1998 (27.03.98)</td> <td>GB</td> </tr> <tr> <td>60/080,325</td> <td>1 April 1998 (01.04.98)</td> <td>US</td> </tr> </table> <p>(71) Applicant (for all designated States except US): <b>UNIVERSITY OF BRISTOL [GB/GB];</b> Senate House, Tyndale Avenue, Bristol BS8 1TH (GB).</p> <p>(72) Inventor; and</p> <p>(75) Inventor/Applicant (for US only): <b>EDWARDS, Keith, Joseph</b> [GB/GB]; IACR-Long Ashton Research Station, University of Bristol, Dept. of Agricultural Sciences, Long Ashton, Bristol BS41 9AF (GB).</p> <p>(74) Agents: <b>SHEARD, Andrew, Gregory et al.; Kilburn &amp; Strode,</b> 20 Red Lion Street, London WC1R 4PJ (GB).</p> </div> <div style="width: 48%;"> <p>(81) Designated States: <b>AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW,</b> ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p> <p><b>Published</b></p> <p><i>With international search report.</i></p> <p><i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p> </div> </div>			9714005.7	2 July 1997 (02.07.97)	GB	9806702.8	27 March 1998 (27.03.98)	GB	60/080,325	1 April 1998 (01.04.98)	US																																																																											
9714005.7	2 July 1997 (02.07.97)	GB																																																																																				
9806702.8	27 March 1998 (27.03.98)	GB																																																																																				
60/080,325	1 April 1998 (01.04.98)	US																																																																																				
<p>(54) Title: <b>METHOD OF DETERMINING THE GENOTYPE OF AN ORGANISM USING AN ALLELE SPECIFIC OLIGONUCLEOTIDE PROBE WHICH HYBRIDISES TO MICROSATELLITE FLANKING SEQUENCES</b></p>																																																																																						
<div style="display: flex; align-items: flex-start;"> <div style="flex: 1; text-align: center;">  </div> <div style="flex: 2; margin-left: 20px;"> <table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left;">NUMBER</th> <th style="text-align: left;">NAME</th> <th style="text-align: left;">SEQUENCE</th> </tr> </thead> <tbody> <tr> <td>1.</td> <td>M017.1</td> <td>5'ACGTCCCCCTCCGG</td> </tr> <tr> <td>2.</td> <td>M017.2</td> <td>5'TACCGAATCCGTAG</td> </tr> <tr> <td>3.</td> <td>OH43.1</td> <td>5'ACGTCCCTCTCCGT</td> </tr> <tr> <td>4.</td> <td>OH43.2</td> <td>5'GGGTGGGGGGGCAA</td> </tr> <tr> <td>5.</td> <td>C0125.1</td> <td>5'GGCCGGGGGGCAATG</td> </tr> <tr> <td>6.</td> <td>C0159.1</td> <td>5'CGCGCGAATCCGTAG</td> </tr> <tr> <td>7.</td> <td>CONTROL</td> <td>5'TCGTTCGGTCCATGAA</td> </tr> </tbody> </table> </div> </div> <div style="margin-top: 10px; text-align: center;"> <table style="margin: auto; border: none;"> <tr> <td>T</td><td>C</td><td>T</td><td>C</td><td>B</td><td>B</td><td>B</td><td>F</td><td>M</td><td>O</td><td>C</td><td>F</td> </tr> <tr> <td>2</td><td>M</td><td>3</td><td>0</td><td>1</td><td>3</td><td>7</td><td>2</td><td>0</td><td>H</td><td>0</td><td>7</td> </tr> <tr> <td>3</td><td>3</td><td>0</td><td>1</td><td>4</td><td>7</td><td>3</td><td></td><td>1</td><td>4</td><td>1</td><td></td> </tr> <tr> <td>2</td><td>7</td><td>3</td><td>5</td><td></td><td></td><td></td><td></td><td>7</td><td>3</td><td>2</td><td></td> </tr> <tr> <td></td><td></td><td></td><td>9</td><td></td><td></td><td></td><td></td><td></td><td></td><td>5</td><td></td> </tr> </table> </div>			NUMBER	NAME	SEQUENCE	1.	M017.1	5'ACGTCCCCCTCCGG	2.	M017.2	5'TACCGAATCCGTAG	3.	OH43.1	5'ACGTCCCTCTCCGT	4.	OH43.2	5'GGGTGGGGGGGCAA	5.	C0125.1	5'GGCCGGGGGGCAATG	6.	C0159.1	5'CGCGCGAATCCGTAG	7.	CONTROL	5'TCGTTCGGTCCATGAA	T	C	T	C	B	B	B	F	M	O	C	F	2	M	3	0	1	3	7	2	0	H	0	7	3	3	0	1	4	7	3		1	4	1		2	7	3	5					7	3	2					9							5	
NUMBER	NAME	SEQUENCE																																																																																				
1.	M017.1	5'ACGTCCCCCTCCGG																																																																																				
2.	M017.2	5'TACCGAATCCGTAG																																																																																				
3.	OH43.1	5'ACGTCCCTCTCCGT																																																																																				
4.	OH43.2	5'GGGTGGGGGGGCAA																																																																																				
5.	C0125.1	5'GGCCGGGGGGCAATG																																																																																				
6.	C0159.1	5'CGCGCGAATCCGTAG																																																																																				
7.	CONTROL	5'TCGTTCGGTCCATGAA																																																																																				
T	C	T	C	B	B	B	F	M	O	C	F																																																																											
2	M	3	0	1	3	7	2	0	H	0	7																																																																											
3	3	0	1	4	7	3		1	4	1																																																																												
2	7	3	5					7	3	2																																																																												
			9							5																																																																												
<p>(57) Abstract</p> <p>A method is described for the determination of the genotype of an organism which comprises the step of hybridising an Allele Specific Oligonucleotide (ASO) probe to a target DNA sequence in which the target DNA sequence comprises a sequence of DNA which flanks a microsatellite repeat unit or Simple Sequence Repeat (SSR).</p>																																																																																						

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon	KR	Republic of Korea	PL	Poland		
CN	China	KZ	Kazakhstan	PT	Portugal		
CU	Cuba	LC	Saint Lucia	RO	Romania		
CZ	Czech Republic	LI	Liechtenstein	RU	Russian Federation		
DE	Germany	LK	Sri Lanka	SD	Sudan		
DK	Denmark	LR	Liberia	SE	Sweden		
EE	Estonia			SG	Singapore		

METHOD OF DETERMINING THE GENOTYPE OF AN ORGANISM USING AN ALLELE SPECIFIC OLIGONUCLEOTIDE PROBE WHICH HYBRIDISES TO MICROSATELLITE FLANKING SEQUENCES

5 The present invention relates to a method of determining the genotype of an organism using an allele specific oligonucleotide (ASO) probe and to a method of designing an ASO probe.

10 Marker assisted breeding relies upon polymorphic genetic markers. Such markers include restriction fragment length polymorphisms (RFLPs) (Edwards, M.D. *et al.*, *Theor. Appl. Genet.* 83, 765-774 (1992)) amplified fragment length polymorphisms (AFLPs) (Vos, P *et al.*, *Nucleic Acids Research* 23, 4407-4414 (1995)) and simple sequence repeats (SSRs) or microsatellite loci (Weber, J.L. and May, P.E., *American Journal of Human Genetics* 44, 388-396 (1989)). The methods for detecting these polymorphic markers all rely upon electrophoretic separation of DNA in inert gels (agarose, acrylamide). For example, at microsatellite loci, the variation in allele lengths arising from differences in the number of repeat units, can be detected by a combination of polymerase chain reaction (PCR) amplification and polyacrylamide gel electrophoresis. Developments in fluorescent DNA fragment analysis not only make it possible to analyse many SSR loci simultaneously but also to automatically capture the data electronically (Ziegle, J. *et al.*, *Genomics* 14, 1026-1031 (1992)). Despite the advent of these semi-automated systems or refinements such as capillary gel electrophoresis, gel-based technology is very labour intensive and time consuming for the large scale genotyping required both in experimental genome analysis and in marker assisted breeding programmes.

25 Microsatellite markers (or Simple Sequence Repeats; CA, CT, AT, etc) are currently a favoured marker for genotyping. They are single locus, co-dominant and multi-allelic and they are based upon the PCR which is relatively cheap to perform and can be automated. Current technology relies on the variability that exists within

the simple sequence repeat copy number, i.e. one genotype may have 20 CT repeats at a specific loci whilst another may have 21 CT repeats. This difference can be detected by gel electrophoresis of the PCR products generated by using PCR primers which flank the simple sequence repeat. This is the main problem with SSRs; gel  
5 electrophoresis is time consuming and expensive to perform. The ideal genotyping test would not employ gel electrophoresis.

The requirements for a high throughput system might include increased scope for automation and a simple binary scoring system that can be reliably read by machine  
10 with no human intervention. Hybridisation between complementary DNA strands which underpins so-called "DNA chips" (Lipshutz *et al.*, 1995) could provide the basis for such a system. In order for hybridisation to be useful for genotyping it is necessary to be able to discriminate between alleles. Such discrimination is possible though the use of Allele Specific Oligonucleotides (ASOs). ASO technology is based  
15 upon the principle that when hybridised under appropriate conditions, synthetic DNA oligonucleotide probes (15-21 bases) will anneal to their complementary PCR generated target sequences only if they are perfectly matched. Under the correct conditions a single base pair mismatch is sufficient to prevent the formation of a stable probe-target duplex.

20

Allele Specific Oligonucleotides (ASOs) have been described previously (Conner *et al Proc. Natl. Acad. Sci. USA* 80 278-282 (1983)). Examples of the application of ASOs have been also reported for detection of genetic disorders in humans (Ala-  
kokko *et al Proc. Natl. Acad. Sci. USA* 87 6565-6568 (1990)); Studenicki *et al DNA*  
25 3 7-15 (1984)) and in characterisation of resistance to fungicides in plant pathogenic fungi (Koenradt, H., & Jones, A. L., *Phytopathology* 82 1354-1358 (1992)). Several tests based on the Allele Specific Oligonucleotide assay are currently used. This assay does not require gel electrophoresis, but is based upon a simply

hybridisation test. However, one major problem with ASOs is finding sufficient polymorphism between the different genotypes at a chosen loci. This problem has meant that ASOs are currently only used for important human disease loci which are of high value.

5

Since ASOs have the potential to be a quick, cheap, multiallelic and multi-locus test, they should be in regular use within genotyping laboratories. Unfortunately, whilst they are in regular use for the detection of certain human genetic diseases, they are not in regular use for non-human genotyping. The reason for this becomes apparent when one considers the enormous cost of developing ASOs. For each locus, a mapped single copy probe has to be sequenced and suitable PCR primers designed. These primer must then be used to amplify the corresponding fragment from all the other possible genotypes. These fragments must then sequenced and the sequences compared with one another to determine ASOs for each of the possible alleles. In addition, when one considers that in an average plant genotyping laboratory, 100 different loci might routinely be screened, then the amount of work required to develop ASOs for each locus and each possible allele, becomes considerable.

The amount of work required to produce ASOs would be considerably reduced if existing molecular markers could be used. These markers would already have been mapped and therefore could be chosen based upon their known and useful map position. Current RFLP markers could therefore offer such a short cut; unfortunately, in a recent study of maize RFLPs it was found that the sequence variation present between different "alleles" was insufficient to design ASOs for all but a few loci. Further, recent unpublished work suggests that microsatellites could provide the basis for an ASO genotyping system; it has been known for a long time that microsatellites are highly polymorphic, presumably due to strand slippage during DNA replication. Unfortunately, this variation (in the repeat unit) is of no

25

20

use in the design of allele specific oligonucleotides (ASOs).

5 A recurrent problem with microsatellite markers in pigs and other animals is the incidence of so-called "null" or pseudo-null alleles. These null alleles are revealed as repeated failure to produce a PCR product and are attributed to polymorphisms at the primer binding site preventing primer annealing. In a recent paper by Alexander *et al.*, (*Animal Genetics* 27, 137-148 (1996)) 50 out of 400 porcine microsatellites were recorded as exhibiting null alleles. In order for sequence variation at the primer binding site to disrupt PCR amplification the variant nucleotide is likely to be located within 5 nucleotides of the 3' end of the primer. Thus, each primer pair effectively assays 10 nucleotide positions for polymorphisms. It is possible to predict the frequency of polymorphisms from the observations of Alexander *et al.*, (*Animal Genetics* 27, 137-148 (1996)) as 50 per 400 x 10 nucleotides i.e. 1 per 80 nucleotides which is 3 to 4 times the level expected in random mammalian genomic DNA.

15 However, it has now surprisingly been found that the sequences flanking the microsatellite repeat unit are also more variable than other single copy DNA. It is believed that this discovery will make the search of ASOs much easier and will result in assay based on Allele Specific Oligonucleotides being much cheaper to set up for most areas of any genome. An objective of the present invention, therefore, is to make the process of mass producing ASOs more efficient and to provide means for use of the ASOs in the analysis of polymorphisms in the genome of an organism.

25 The reasons for the enhanced level of variation in the sequences flanking a microsatellite repeat unit are unclear, but it is presumably due to the close proximity of the simple sequence repeat. When one considers that for all microsatellite markers, primers have already been made, the marker has been mapped, and the

amplified fragment are sufficiently small to be sequenced in one single run on an ABI 377, then previously characterised microsatellite markers may become the markers of choice for developing large numbers of ASOs. From the existence of "null" or pseudo-null alleles in animals, this further suggests that there may also be greater levels of sequence variation around microsatellite loci in animals.

According to a first aspect of the present invention there is provided a method of determining the genotype of an organism, the method comprising the step of hybridising an Allele Specific Oligonucleotide (ASO) probe to a target DNA sequence, wherein the target DNA sequence comprises a sequence of DNA which flanks a microsatellite repeat unit or Simple Sequence Repeat (SSR).

The genotype of an organism can be defined as the genetic constitution of an individual organism, as distinct from its phenotype which is the total appearance of an organism determined by interaction during development between its genetic constitution and the environment. Different phenotypes may result from identical genotypes, but is generally unlikely that two organisms could share all their phenotypic characters without having identical genotypes.

The present invention is applicable to all organisms, particularly plants, animals and fungi, including yeasts. Methods of the present invention may find utility to species of non-flowering and flowering plants, both monocotyledonous and dicotyledonous. Plant species of interest include, but are not limited to maize (*Zea mays*), teosinte, *Arabidopsis thaliana*, Brassica spp., cereals (e.g. oats, barley, wheat, rye), banana, palms, ornamental plants (e.g. orchids, lilies, tulips, roses, clematis), trees (e.g. forest, fruit or ornamental trees), shrubs, tobacco, potatoes, beans, yams, cassava, sunflower, tomato, pepper, cucumber, lettuce or rice. Yeast species include, but are not limited to, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pischia*,

*Kluyveromyces lactis*. Fungal species include, but are not limited to both pathogenic and non-pathogenic fungi, for example the wheat fungal pathogen *Septoria*. In principle, the invention is also applicable to all animals, including birds, such as domestic fowl, amphibian species, reptile species and fish species. In practice, however, it will be to animals, especially marsupials or mammals, particularly placental mammals that the greatest commercially useful application is presently envisaged. It may find general application to humans and also to non-human animals, preferably mammals. It is with ungulates such as cattle, sheep, goats, water buffalo, camels and pigs that the invention is likely to most useful. It should also be noted that the invention is likely to be applicable to other economically important animal species such as, for example, horses, llamas, alpacas or rodents, e.g. mice, rats or rodents.

The invention is equally applicable to the determination of the genotype of a transgenic organism as defined above which is prepared by methods known in the art, including recombinant DNA technology, DNA/RNA transfection procedures, nuclear transfer technology ("cloning") and DNA/RNA microinjection. Mass transfection or transformation techniques can also be used, e.g. electroporation, viral transfection or lipofection, suitably with liposome delivery. It should be noted that the term "transgenic" should not be taken to be limited to referring to an organism as defined above containing in their germ line one or more genes from another species, although many such organisms will contain such a gene or genes. Rather, the term refers more broadly to any organism whose germ line has been the subject of technical intervention by recombinant DNA technology. So, for example, an organism in whose germ line an endogenous gene has been deleted, duplicated, activated or modified is a transgenic organism for the purposes of this invention as much as an organism to whose germ line an exogenous DNA sequence has been added.



Hybridisation of an oligonucleotide probe to a target DNA sequence in a method according to the present invention, includes the denaturation of a source of duplex DNA, also known as "melting", in which the DNA is allowed to anneal with an appropriate oligonucleotide probe under certain conditions. Denaturation of double-stranded DNA (dsDNA) may be achieved by small increases in the temperature of DNA in solution. Alternative means include the use of variations in pH or the use of chemical agents such as urea, alcohols or detergents. Annealing of DNA with an oligonucleotide probe after "melting" can be achieved by the reversal of the denaturation condition, e.g. by a small decrease in the solution temperature. However, in methods according to the present invention, a reasonable degree of specificity of hybridisation is desired and so relatively stringent conditions may be used to form the duplexes of probe and DNA sequence to be amplified. Such stringent conditions may be characterised by low salt concentration or high temperature conditions.

As used in the present application, the term "highly stringent conditions" means hybridisation to DNA bound to a solid support in 0.5M NaHPO<sub>4</sub>, 7% sodium dodecyl sulfate (SDS), 1mM EDTA at 65°C, and washing in 0.1xSSC/0.1% SDS at 68°C (Ausubel *et al* eds. "Current Protocols in Molecular Biology" 1, page 2.10.3, published by Green Publishing Associates, Inc. and John Wiley & Sons, Inc. New York (1989)). In some circumstances, less stringent hybridisation conditions may be required. As used in the present application, the term "moderately stringent conditions" means washing in 0.2xSSC/0.1% SDS at 42°C (Ausubel *et al* (1989) *supra*). Hybridisation conditions can also be rendered more stringent by the addition of increasing amounts of formamide, to destabilise the hybrid duplex. Thus particular hybridisation conditions can be readily be manipulated, and will be generally be selected according to the desired results. In general, convenient

hybridisation temperatures in the presence of 50% formamide are: 42°C for a probe which is 95 to 100% homologous to the target DNA, 37°C for 90 to 95% homology, and 32°C for 70 to 90% homology.

5 The complexity of the hybridisation reaction carried out depends upon the length of the sequence on the array and the application (reviewed in Marshall, A. and Hodgson, J., *Nature Biotechnology* **16** 27-31 (1998)). For polymorphism analysis or resequencing, where every nucleotide position in a gene exon or mutation hot spot has to be interrogated, a set of four oligonucleotides is generally designed (one for  
10 each base type) that spans each position in the target sequence, differing only in the identity of the central base. The relative intensity of hybridisation to each series of probes at a particular location identifies the base. Each set of oligonucleotides is offset by one base so that they can be arranged in order by analysing overlaps, a process known as "tiling". If the application is expression monitoring, where details  
15 of the precise sequence are unnecessary, sets of oligonucleotides are constructed that identify unique motifs in genes. By arranging them in a particular order, it is possible to identify chromosomal location as well as sequence. In contrast, arrays of cDNA work more like conventional dot-blot where competitive hybridisation of two labelled samples (disease versus normal; heat-shock induced versus normal) reveals  
20 different gene expression.

Allele specific oligonucleotides (ASOs) in accordance with the present invention may comprise oligonucleotide or short polynucleotide sequences of deoxyribonucleotides known as "bases" which typically include one or more of deoxyriboadenosine (A),  
25 deoxyribocytosine (C), deoxyriboguanosine (G) and deoxyribothymosine (T). Other possible component bases include deoxyriboinosine (I) and/or chemically-modified variants A, C, G, T, or I, for example methylated derivatives. The ASO probe sequence may be from 10 to 50 bases long, suitably 12 to 15 bases, preferably 15 to

21 bases. The precise length can be selected based upon the target DNA sequence to be hybridised with by the ASO. Alternatively, the ASO probe may comprise a peptide nucleic acid (PNA) (Nielsen *et al Science* 254 1497- (1991); Eghom *et al J. Am. Chem. Soc.* 114 1895- (1992); Llanvey *et al Science* 258 1481- (1992)). PNAs are nucleic acid analogues composed of a polymer of 2-aminoethyl glycine which acts as the backbone of the molecule. Each monomer is linked by a methylenecarbonyl linkage to one of the bases found in DNA or RNA. The use of a PNA molecule as the ASO has the effect of increasing the stability of the correct hybridisation versus the mismatch hybridisation.

In certain situations it may be beneficial to be able to enhance the discrimination of single nucleotide polymorphisms by artificial mismatch hybridisation as described by Guo *et al (Nature Biotechnology* 15 331-335 (1997) and this technique can be applied in methods according to the present invention. Artificial mismatches can be inserted into an ASO probe using the base analogue 3-nitropyrrole or another equivalent molecule. A significant enhancement of the discrimination can therefore generally be obtained in this way if desired. These modifications to the ASO have the effect of decreasing the stability of the mismatch hybridisation more than the match hybridisation.

The target DNA sequence is a short region of DNA in the DNA of the organism to be investigated, in which the target sequence flanks a microsatellite repeat unit. The DNA sequence may be an oligonucleotide or a polynucleotide sequence as appropriate, of from 5 to 500 base pairs, suitably of from 10 to 100 base pairs, preferably, 15 to 45 base pairs. A microsatellite repeat unit or Simple Sequence Repeat (SSR) is a marker in a DNA sequence which is single locus, co-dominant and multi-allelic. SSRs are short tandem repetitive DNA sequences with a repeat length of a few (1 to 5) base pairs. For example, the repeat unit may be any base, e.g., A,

G, C, or T, or any pair of bases, e.g. CA, CT, AT, etc. The SSR may be repeated up to around 8 to 10 repeat units roughly every 10-20 kilobases in the genome of an organism. Higher number repeat unit units of up to 100 repeats are often called "minisatellite" DNA regions and beyond 100 repeat units the terminology is generally "satellite" DNA regions.

Methods in accordance with the present invention to determine the genotype of an organism can also suitably make use of the Polymerase Chain Reaction (PCR) to identify the hybridisation of an ASO to a region flanking a microsatellite region of interest. PCR can be carried out, e.g. by use of a Perkin-Elmer / Cetus thermal cyclor with *Taq* polymerase (Gene Amp<sup>TM</sup>) or *Taq* Gold polymerase as described in Erlich *et al* (*Nature* 331 461-462 (1988)).

An extension of approaches using the PCR method is the use of allele-specific amplification (ASA) and such techniques may also be suitably employed in methods according to the present invention. Allele-specific amplification (ASA) is the basis for a number of rapid, reliable, non-isotopic techniques, which depend on selective PCR amplification (reviewed by Prosser, J. in *TIBTECH* 11 238-246 (1993)). In ASA, several DNA regions (i.e. several mutations) can be amplified in one reaction (multiplex analysis). Some oligonucleotides carry the mutation in the centre of the molecule so that differential amplification depends on differential hybridisation (competitive oligonucleotide priming or COP). More often the method depends upon placing the mutation at the extreme 3'-end of one primer where, under appropriate conditions, mismatch can prevent, or severely reduce polymerase extension, in other words, amplification (abbreviated as amplification-refractory mutation system or ARMS). Further modifications of this technique are known in the art and can also be utilised within the present invention if appropriate.

Allele specific oligonucleotide hybridisation can be used to identify any known mutation and involves differential hybridisation of sequence-specific oligonucleotides. The oligonucleotides can be suitably prepared with the mutation placed centrally and are preferably hybridised to target DNA under conditions which permit hybridisation only if a perfect match is found. If the oligonucleotides are hybridised to PCR-amplified target DNA (dot-blot technique), one mutation can be tested per reaction (Saiki *et al Nature* 324 163-166 (1986)), but when the allele specific oligonucleotides are attached to the hybridising membrane and hybridised with labelled target DNA (reverse dot-blot), a number of different mutations in one fragment can be tested (Saiki *et al Proc. Natl. Acad Sci. USA* 86 6230-6234 (1989)).

By way of illustration and summary, the following scheme sets out a typical process by which the genotype of an organism can be determined according to a method of the present invention. The process can be regarded as involving five steps:

- (1) selection of microsatellite marker in organism whose genotype is to be determined;
- (2) generation of target DNA sequences by PCR using appropriate primers;
- (3) comparison of sequence to design ASO probes;
- (4) synthesis of ASO probes;
- (5) hybridisation of ASO probe and PCR-amplified target DNA sequence; and
- (6) analysis of results.

Microsatellite markers can be selected from available sources of genome information. Selection is preferably on the basis of the microsatellite motif (i.e. if

there are differences), map location (such that they are evenly spaced throughout the genome) and the ability of their PCR primers to generate product.

Amplification of the corresponding target DNA sequences by PCR can be carried out conveniently using any available PCR protocol (McPherson eds. *et al* in "*PCR: A Practical Approach*", IRL Press, Oxford England (1991); McPherson, M. J. and Hames, B. D., eds. in "*PCR 2: A Practical Approach*", IRL Press (1995); B. White ed. in "*PCR Cloning Protocols*" *Methods in Molecular Biology* 67 (1996); or Dieffenbach, C. and Cveksler, G. S., eds. in "*PCR Primer: A Laboratory Manual*", Cold Spring Harbor Laboratory (1995)).

Synthesis of ASO probes can be achieved by standard chemical synthetic routes in the art comprising ligating together successive nucleotides and/or oligonucleotides. ASO probes can also be prepared using reverse transcriptase to transcribe a desired RNA sequence in which the case the oligonucleotide will be a cDNA molecule.

Hybridisation of ASO probes to target DNA sequences may be performed using the dot-blot or reverse dot-blot techniques described above. Alternatively, the ASO probe can be bound to a solid support which can be any suitable inert material, e.g. nitrocellulose, polyethylene, polypropylene, silicon (Lipshutz, R.J. *et al*, *Biotechniques* 19 442-447 (1995)), or glass. A glass support with DNA probes bound to the surface can be termed a "DNA chip" or "oligonucleotide chip" and methods which use such embodiments are also within the scope of the present invention. The production of DNA chips is comprehensively reviewed in Mirzabekov, A. in *Trends Biotechnol.* 12 27-32 (1994) and further described in Shalon *et al* *Genome Res.* 6 639-645 (1996); Drobyshchev *et al* *Gene* 188 45-52 (1997); Marshall, A., and Hodgson, J., *Nature Biotechnology* 16 27-31 (1998)).

The ASO probe or probes bound to the surface of the inert support can be suitably arranged in the form of an array. At present, the several means by which an array can be made fall into three general categories: *in situ* (on-chip) synthesis of oligonucleotides or peptide nucleic acids (PNAs); arraying of prefabricated  
5 oligonucleotides/PNAs; and, spotting of DNA fragments. However, the present invention is not limited with respect to the means chosen to prepare the array probes on a solid support.

For *in situ* oligonucleotide synthesis, the array can be prepared by photolithography  
10 or piezoelectric printing. In the photolithographic method, a mercury lamp is shone through a photolithographic mask onto the chip surface, which removes a photoactive group, resulting in a 5'-hydroxyl group capable of reacting with another nucleoside. The mask can be used in this way to determine which nucleotides become activated. Successive rounds of deprotection and chemistry can result in  
15 oligonucleotides which are up to 30 bases in length. The piezoelectric method uses a "printer-head" (analogous to an ink-jet printer head) which travels across the array and at each spot to be applied, a microlitre drop of one of the four main bases (A, G, C, or T) is spotted onto the coated surface where it is anchored by standard chemistry. Following washing and deprotection, the next cycle of oligonucleotide  
20 synthesis is carried out. The method can be used to prepare oligonucleotides of up to 40-50 bases in length which represents an improvement over traditional controlled pore glass (CPG) oligonucleotide synthesis.

The construction of arrays can also be simplified by prefabricating oligonucleotides  
25 or oligopeptides using conventional CPG methods and then by printing them onto the array using direct touch or micropipetting. An alternative to this approach is to employ a controlled electric field to immobilise prefabricated oligonucleotides to spots (microelectrodes) on the array. This method uses biotinylated oligonucleotides

which are directed to individual spots by polarising the charge at the spot and are then anchored at the spot via a streptavidin-containing permeation layer that covers the surface.

5 The third main method of preparing an array is to "spot" oligonucleotides directly onto the inert support surface. For example, glass slides can be overlaid with a positively charged coating, such as amino silane or polylysine, and oligonucleotide fragments suspended in a denaturing solution are then printed directly onto the surface.

10

Analysis of the results of the hybridisation step can be achieved using standard techniques to detect the presence of chemical, fluorescent or radioactive labels attached to the target DNA sequence. Chemical labels can include but are not limited to, biotin, avidin, horseradish peroxidase, alkaline phosphatase, or other  
15 visually detectable coloured dyes. Fluorescent labels can include, but are not limited to, fluorescein, or other optically detectable fluorescent dyes. Radioactive labels can include, but are not limited to,  $^{32}\text{P}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$ ,  $^3\text{H}$ ,  $^{125}\text{I}$ . Alternatively, the analysis of the results can utilise laser desorption to interrogate the hybridisation of probe to target with a readout generated by mass spectrometry. Mass spectrometry can also  
20 be used alone or in conjunction with fluorescence or chemical markers depending upon the level of automation in the procedure.

According to a second aspect of the present invention there is provided the use of a microsatellite repeat unit or Simple Sequence Repeat (SSR) in the design of an allele  
25 specific oligonucleotide (ASO) probe.



According to a third aspect of the present invention there is provided the use of a microsatellite repeat unit or a Simple Sequence Repeat (SSR) primer set to amplify PCR products for use in a method according to the first aspect of the invention.

5 According to a fourth aspect of the present invention there is provided an inert support comprising a plurality of allele specific oligonucleotide probes (ASOs) bound to the surface of the support, in which the ASOs specifically hybridise to a target DNA sequence which flanks a microsatellite repeat unit or Simple Sequence Repeat (SSR). The support may any inert material as described above.

10

According to a fifth aspect of the present invention there is provided a kit for the determination of an organisms genotype comprising an inert support in accordance with the fourth aspect of the invention and a means for detection the hybridisation of an ASO to a target DNA sequence as defined previously,

15

Preferred features and characteristics of the second and subsequent aspects are as for the first aspects *mutatis mutandis*.

20

The invention will now be described by way of example with reference to the accompanying Examples and drawings which are provided for the purposes of illustration and neither of which are to be construed as being limiting on the present invention. In the description of this application, reference is made to the following drawings in which:

25

FIGURE 1 shows an example of maize allele specific oligonucleotides demonstrating hybridisation/non-hybridisation of the probes.

FIGURE 2 shows a comparison of sequences from Maize for microsatellite MACE01F07: Forward Primer.

FIGURE 3 shows a comparison of sequences from Septoria for 312F.

FIGURE 4 shows an alignment of pig SSR sequences for pig SH524.1 and pig SH525.1.

FIGURE 5 shows the conversion of microsatellite flanking sequences to ASOs and their use in genotyping 12 maize lines. The 6 allele specific oligonucleotides (ASOs) and one control oligonucleotide designed from the flanking sequence are shown alongside the results of the genotyping presented as a compilation autoradiograph.

#### Examples 1 to 5 – Plants, specifically maize and fungus

Marker assisted selection is already widespread in plant breeding. The markers used in such plant breeding schemes include RFLPs, AFLPs and SSRs. The main limitations of these gel-based technologies are that they both limit the number of samples that can be characterised to just a few hundred per day and increase the costs of the genetic screening. The genetic material used in plant breeding is based upon extensive collections of semi-characterised lines, which, depending upon the species can vary from genetically identical and homozygous for all markers, to open pollinated material with a genetic structure similar to that found in animal populations. Currently, a number of research groups are developing microsatellite markers for a range of crop species and these are rapidly becoming the marker of choice for high throughput genotyping.

Hybridisation/non-hybridisation of a probe to a selected DNA region can then be monitored via a suitable detection system. This two state system (hybridisation/non-hybridisation) is binary in nature as shown in Figure 1 and is ideal for interpretation by machines. If the single locus PCR product, hybridises to two allelic ASOs then the individual must be heterozygote, where as hybridisation to one ASO suggests that it is homozygote.

**Example 1:**

### Example of variation in sequences flanking microsatellite repeat unit in maize

10 The variation in the sequences flanking the microsatellite repeat units is as shown below, in which 43 bases of sequence have been taken from one side of a single locus maize SSR from 7 different inbred lines:

```

Line 1:TCTCTCTCTCTCTCGTCTCTCTCGGACGCTTGTAACCTCTACT
15 Line 2:TCTCTCTCTCTCTCACTCTCTCTCCTCATATCACCCCCACT
Line 3:TCTCTCTCTCTCTC...TCTCTCGACACGTGTGTCTCTCTCT
Line 4:TCGCTCTCGCCCTCGTCTCTCTCGGACGCTTGTAATCCTACT
Line 5:TCGCTCTCGCCCTCGTCTCTCTCGGACGCTTGTAATCCTACT
Line 6:TCTCTCTCTCTCTC...TCTCTCCACACTTGTAATCCTACT
20 Line 7:TCGCTCTCGCCCTCGTCTCTCTCGGACGCTTGTAACCTCTACT
Line 8:TCGCTCTCGCCCTCGTCTCTCTCGGACGCTTGTAACCTCTACT
Changes . . . . . . . . . .

```

Lines 7 and 8 are the same inbred line.

25

In this case out of 43 bases, 19 bases are variable between the 7 lines compared to approximately one base per 200 bases in randomly chosen RFLP markers.

Figure 5 shows the results of genotyping 12 maize lines as a compilation autoradiograph with the ASOs shown in the Figure alongside their respective autoradiograph. The ASOs were designed from the flanking sequences of the maize sequences shown in Figure 2. The genotyping exercise successfully allowed  
5 determination of all the maize lines except for line T232. However, by using the ASOs it was possible to state that T232 is more closely related to T303, CO159, B14, B37, B73 and F2 than it is to CM37, MO17, OH43, CO125 and F7.

### Materials and Methods

10 PCR products from the amplification of maize DNA (using one of the 12 lines from Figure 5) with the publicly available MACE01F07 SSR primers

Forward primer: 5'-TCGTTCGGTCCATGAAAT

15 Reverse primer: 5'-CAAATATCTCTCATCTTTGCTGAC

were denatured and spotted onto Hybond N<sup>TM</sup> membrane to yield 7 identical strips as indicated in Figure 5. Each one of these strips was hybridised with the appropriate ASO (nos. 1 to 7). Previously, the ASOs had been labelled with  $\gamma$ -<sup>32</sup>P-ATP. The  
20 hybridisation conditions were as follows: 6xSSC, 1% SDS, 0.25% Marvel<sup>TM</sup> milk, 1ng/ml labelled ASO. The temperature of hybridisation was 37°C and the time of hybridisation was 3 hours. Following hybridisation, the strips were washed in 1xSSC, 1% SDS and at a temperature of between 37°C and 60°C depending upon the particular ASO. Following washing, the strips were subjected to autoradiography  
25 for 20 hours. The strips were then brought together for the photograph seen in Figure 5.

### Example 2:

Example of variation in sequences flanking microsatellite repeat unit in Septoria

The variation in the sequences flanking the microsatellite repeat units in Septoria, a fungal pathogen of wheat is as shown in Figure 3.

5     **Example 3:**Selection of Microsatellite Markers

Several hundred maize microsatellite primer sets have been previously characterised (Edwards, K.J. *et al.*, *BioTechniques* 20 (5) 758-760 (1996)). Approximately five microsatellite markers for each of the available microsatellite motifs (CA, CT, CAA, etc) will be examined to evaluate the variability of the flanking sequences in the different SSR types. If there is a difference in the amount of variation in the flanking sequences, within the different motifs, then that specific type will be selected for future work. From the available microsatellites, 96 (mapped and characterised) markers will be selected for inclusion in this study. These will be selected on the basis of their microsatellites motif (if there are differences), map location (such that they are evenly spaced throughout the genome) and the ability of their PCR primers to generate a product from the chosen lines.

Example 4:20     Selection of Lines

15 maize inbred lines have been chosen based upon their heterotic characteristics and their importance to current breeding (European and US) efforts. These lines are currently being used and the results obtained in this study will be compared to existing results.

25

The 15 lines are:     1:A554, 2:B73, 3:CM37, 4:CO159, 5:F2, 6:H99, 7:MO17,  
                          8:OH43, 9:PA91, 10:T232, 11:Tx303, 12:W64a, 13:CO125,  
                          14:F7 and 15:F564.

**Example 5:****Sequencing/ASO Design**

The previously characterised, flanking primer sets for each of the 96 microsatellite markers will be used to amplify and sequence the corresponding loci from the 15 maize inbred lines. For this, a Biomek 2000 robot and a ABI377 automated sequencer will be utilised. Sequences generated will be compared via the programme CLUSTALW and the information generated used to design ASOs for the different genotypes. In the unlikely event of a marker not producing sufficient variation to design ASOs for each genotype then a linked marker will be sequenced in its place. ASOs for both the same locus and the different loci will be designed to have the same annealing temperatures (for an exact match). This process is not an exact science and it is expected that a number of the ASOs will need to be re-designed. In order to help reduce the effect of this likely problem and enhance the discrimination between matched and mismatched ASOs, artificial mismatches using the 3-nitropyrrole base analogue will be included. This can have the effect of increasing the differential hybridisation by as much as 200% (Guo, Z *et al.*, *Nature Biotechnology* 15, 331-335 (1997)).

**Example 6:****Production of the ASO Glass Plate (chip)**

Several methods have been put forward to bind ASOs to a solid support. The method of Guo, *et al.*, ( *Nucl. Acids Res.* 22, 5456-5465 (1994)) will be utilised which is straightforward and can be automated on the Biomek 2000 robot. Briefly, standard oligonucleotides will be immobilised on activated glass (8 x 12 cm) plates prepared as follows: firstly, washed in 1% 3-aminopropyltrimethoxysilane in 95% acetone/water for 2 minutes then 0.2% 1,4 phenylene diisothiocyanate in 10% pyridine/dimethyl formamide for 2 hours and washed with methanol and acetone.

Using the Biomek 2000 up to 1,500 ASOs (96 loci, 15 alleles per locus) will be spotted onto each glass plate. As each plate is 8 x 12 cm, this means that the ASOs that make up each loci can be arranged as a 4x4 array (accommodating up to 16 different ASOs per locus) with up to 96 loci per plate. It is believed that this is the minimum spacing that can be employed without the requirement for a dedicated computer system to score the results. Experiments will be performed to determine the amount of oligonucleotide required for each detection. The current work in screening 8 x 12 cm arrays of YAC DNA means that the hybridisation conditions for these plates have already been optimised.

#### Example 7:

##### Testing

The initial studies will employ glass chips containing ASOs from only 10 loci. These chips will be used in hybridisation studies which include the appropriate denatured PCR products from 1 to 10 different PCR reactions, including known controls. Products will be amplified from their respective loci using the original microsatellite PCR primers. During this initial phase of the work, hybridisation of PCR products to their respective ASOs will be detected via the incorporation of biotin into the product during the PCR amplification. The presence of the biotin will be detected via a commercially available kit based upon alkaline phosphatase.

As part of the current studies, procedures are being developed to multiplex microsatellite amplification reactions. These studies have shown that PCR primer design and primer concentration is critical to multiplex PCR. Considerable success using "Taq Gold" in multiplexed reactions has been achieved. The enzyme activity of Taq Gold is negligible until the enzyme has been heated for approximately 10 minutes at 90°C. Hence it allows the operator to perform a very stringent hot start PCR which has the effect of reducing the number of early, false primings, this in

turn leads to a considerable improvement in the specificity of the multiplexed PCR. This experience will be utilised to develop the 96 PCRs (representing 96 loci) into 10 multiplexed reactions. The data obtained from these initial studies will be directly compared with the information derived from our existing microsatellite programme.

5

The complexity of the ASO plate will be increased via the use of fluorescent dyes incorporated into the PCR products. Along with the use of suitable detection equipment, this should increase the number of loci per plate by at least a factor of four or it could be used to increase the number of individual plants screened per plate from one to four. Another alternative approach will be to develop a silicon microchip version of the plate via photolithography based technology (Lipshutz *et al.*, 1995).

10

#### Examples 8 to 10: Animals, specifically pigs

15 Selective animal breeding is currently based on selection indices that take account of multiple selection objectives (i.e. selection for more than one trait at a time). Each trait of economic importance is likely to be influenced by several quantitative trait loci (QTL). To compete with traditional selective breeding practices, marker assisted selection will either have to deliver very large gains by focussing on major genes or else the ability to scan genomes for multiple QTL. To date marker assisted selection in animals has been restricted to a few major genes – for example, the ‘Halothane’ gene which causes porcine stress syndrome, ESR as a predictor of litter size in pigs, the mutation which causes bovine leukocyte adhesion deficiency GLAD. The current markers of choice for genome scanning in vertebrates (humans mice and 20 farmed animals) are so-called microsatellites or simple sequence repeat (SSR) loci. These markers are abundant and evenly distributed throughout the genome (e.g. pigs – Winter, A.-K. *et al.*, *Genomics* 12, 281-288 (1992)). For each of the major farmed animal species – cattle, pigs, sheep and chickens 500-1500 microsatellite 25



markers have been characterised and mapped – [see livestock genome databases mounted or accessible from the Roslin Institute – [http://www.ri.bbsrc.ac.uk/genome\\_mapping.html](http://www.ri.bbsrc.ac.uk/genome_mapping.html)]. The farmed animals used for selective breeding are outbred and relatively heterogeneous. Although animals are classified into breeds or lines, these groupings are not true breeding lines in the sense understood by a plant breeder or a mouse geneticist working with inbred lines. There is genetic variation within breeds and lines so that for highly polymorphic markers such as microsatellites one would expect to find a high frequency of heterozygous individuals.

There are more than a thousand microsatellite loci mapped in the pig. Information on most of these markers is freely accessible in the pig genome database (PiGBASE – [http://www.ri.bbsrc.ac.uk/pigmap/pig\\_genome\\_mapping.html](http://www.ri.bbsrc.ac.uk/pigmap/pig_genome_mapping.html)). Work will focus on evaluating the extent of sequence variation flanking microsatellite repeats and developing multiplex PCR for set of pig microsatellite markers.

15

**Example 8:**

**Selection of microsatellites in pigs**

Ninety-six mapped microsatellite loci selected to give good coverage of the pig genome will be amplified and sequenced in a range of experimental and commercial pigs. These pigs will include four Chinese Meishan and four Large White pigs which constitute the grandparents of a QTL-mapping population at Roslin and which are also the grandparents of the Roslin contribution to the PiGMAP international reference mapping pedigrees (Archibald, A.L. *et al.*, 1995, *Mammalian Genome* 6, 157-175 (1995)). The two Wild Boars used in the Uppsala QTL-mapping populations and also part of the PiGMAP pedigrees (Andersson, L. *et al.*, *Science* 263, 1771-1774 (1994); Archibald, A.L. *et al.*, *Mammalian Genome* 6, 157-175 (1995)) will also be analysed. These ten pigs from three breeds are key individuals in major genome mapping experiments. These animals are also the founders of

extensive structured experimental pedigrees in which polymorphic markers could be mapped at a future date. Pigs from two rare breeds – the Tamworth and Middle White will also be samples. The remaining pigs will be selected from commercial populations. The primary reason for developing this new genotyping technology is to allow marker assisted selection in commercial populations. Thus, the markers need to be sufficiently polymorphic and informative in commercial lines of pigs. Four pigs will be sampled from two major pig breeders – Pig Improvement Company and Cotswold Pig Development Company Ltd. A sample of four pigs is large enough to have a 85% chance of detecting alleles with a frequency of 0.2 (Skolnick, N.H. and White, R., *Cytogenetics and Cell Genetics* 32, 58-67 (1982)).

Figure 4 shows an alignment of pig SSR sequences for pig SH524.1 and pig SH525.1 with differences marked by a “\*” underneath the appropriate residue.

**Example 9:**

Sequence analysis of variation

The microsatellite loci (n=96) will be amplified by PCR in a sample of 20 pigs as described above. The lengths of the PCR products will be determined for each marker and animal on an ABI 373 DNA sequencer with ABI Genescan software. As it is likely that some or most of the pigs sampled will be heterozygous for length variants at these microsatellite loci it will be necessary to isolate the allelic products by gel purification. Where the allelic products are sufficiently different in size gel purification will be effected by high resolution agarose gel electrophoresis in Metaphor agarose. Where the allelic products are of similar size it will be necessary to use acrylamide gels. PCR amplification of dinucleotide repeats such as [dCdA]<sub>n</sub> which is the most abundant porcine microsatellite motif yields not only a fragment corresponding to the sequence as found in the template genomic DNA, but also minor artefactual products one and two repeat units smaller than the authentic

product i.e. two and four nucleotides shorter. The proportion of these 'stutter bands' relative to the authentic product varies from one microsatellite locus to the next. It has been demonstrated that these stutter or shadow bands arise solely from variation in the length of the dinucleotide repeats (Hauge, X.Y. and Litt, M., *Nucl. Acids Res.* 2, 411-415 (1993)). The sequence of the DNA flanking the tandem repeats is faithfully reproduced during the PCR amplification. Each microsatellite allele will be sequenced from both ends but only information from the primer to the start of the repeats will be gathered. Thus, the potential difficulties of interpreting sequence data from the authentic PCR product and the stutter bands which are out of register will be avoided. The PCR products/alleles will be sequenced using cycle sequencing protocols and analysis on an ABI 373 DNA sequencer. Sequence analysis and comparison of the sequences from different alleles will be performed as described above for plants.

**Example 10:**

**Multiplex PCR**

The current gel-based technology makes fewer demands for simultaneous genotyping of multiple markers as the gel analysis, which has a capacity for about 9 markers per individual per gel, is the rate limiting step. Therefore there have only been limited efforts to multiplex the PCR step of the genotyping of pig microsatellites – at present most markers are amplified independently and then the PCR products for up to 9 loci are pooled and analysed simultaneously on the ABI sequencer. The DNA chip/ASO approach could require 96 or more markers to be analysed simultaneously for each individual. The key number for automation in biology laboratories is 96 (8 x 12 matrices). Thus, marker sets will be developed for multiplex PCR in groups of 8. The development of the conditions for multiplex PCR is independent of the viability of ASO technology in animals as determined from the results of the experiments described above. If there is insufficient variation flanking microsatellite loci in pigs

to sustain an ASO/DNA chip approach, then the multiplex PCR methods could be applied to current gel-based genotyping systems.

#### Schedule

- 5     1.     Develop 8 marker PCR multiplex.
2.     Determine allele fragment sizes for these 8 markers in sample of 20 pigs.
3.     Amplify one marker for all 20 pigs.
4.     Gel purify allelic fragments (20 pigs, 1 marker, i.e. up to 40 fragments from  
             acrylamide gels).
- 10    5.     Cycle sequencing chemistry, load gel (up to 40 fragments).
6.     Sequence analysis, data handling.

Once the first multiplex set is established, PCR optimisation for multiplexes will be conducted simultaneously with sequence analysis of marker sets optimised earlier.

- 15    The rate limiting step in the procedure will be the purification of the allelic PCR products from gels. If the requirement for gel purification can be restricted to agarose gels, the timetable for sequencing the allelic variation for the 96 selected markers can be condensed. Any saving in time arising in this manner will be used to
- 20    sequence a larger sample of pigs in order to increase the probability of detecting more alleles. Alternatively, a comparison between DNA flanking microsatellite and other locations could be effected by analysis of randomly selected genomic sequences of similar size. Non microsatellite loci would be isolated at random from small fragment genomic libraries, sequenced, PCR primers developed and the loci sequenced in all 20 pigs as outlined above. As the difficulties arising from the
- 25    length variation and stutter bands inherent to microsatellite loci, these random control loci could be sequenced without recourse to gel purification and thus more quickly. These random loci would be assigned to chromosomes using the INRA

Toulouse somatic hybrid cell panel (Robic, A. *et al.*, *Mammalian Genome* 7, 438-445 (1996)).

CLAIMS

1. A method of determining the genotype of an organism, the method comprising the step of hybridising an Allele Specific Oligonucleotide (ASO) probe to  
5 a target DNA sequence, wherein the target DNA sequence comprises a sequence of DNA which flanks a microsatellite repeat unit or Simple Sequence Repeat (SSR).
2. A method as claimed in claim 1, in which the organism is a plant, animal, yeast or fungal species.  
10
3. A method as claimed in claim 2, in which the plant species is *Zea Mays*.
4. A method as claimed in claim 2, in which the fungal species is *Septoria*.
- 15 5. A method as claimed in claim 2, in which the animal is a mammal.
6. A method as claimed in claim 5, in which the animal is a human.
7. A method as claimed in claim 5, in which the animal is an ungulate.  
20
8. A method as claimed in claim 2, in which the ungulate is selected from cattle, sheep, goats, water buffalo, camels or pigs.
9. A method as claimed in any one of claims 1 to 8, in which the allele specific  
25 oligonucleotide probe sequence is from 10 to 35 bases long, suitably 12 to 15 bases, preferably 15 to 21 bases.

10. A method of determining the genotype of an organism, the method comprising the step of hybridising an Allele Specific Oligonucleotide (ASO) probe to a target DNA sequence, wherein the target DNA sequence comprises a sequence of DNA which flanks a microsatellite repeat unit or Simple Sequence Repeat (SSR) in which the method comprises the following steps:

- (1) selection of microsatellite marker in organism whose genotype is to be determined;
- (2) generation of target DNA sequences by PCR using appropriate primers;
- (3) comparison of sequence to design ASO probes;
- (4) synthesis of ASO probes;
- (5) hybridisation of ASO probe and PCR-amplified target DNA sequence; and
- (6) analysis of results.

11. A method as claimed in claim 10, in which the ASO probe is bound to an inert support.

12. A method as claimed in claim 10, in which the analysis of the results is performed by detecting a chemical, fluorescent or radioactive label attached to the target DNA.

13. The use of a microsatellite repeat unit or Simple Sequence Repeat (SSR) in the design of an allele specific oligonucleotide (ASO) probe.

14. The use of a microsatellite repeat unit or a Simple Sequence Repeat (SSR) primer set to amplify PCR products for use in a method as defined in any one of claims 1 to 12.

15. An inert support comprising a plurality of allele specific oligonucleotide probes (ASOs) bound to the surface of the support, in which the ASOs specifically hybridise to a target DNA sequence which flanks a microsatellite repeat unit or  
5 Simple Sequence Repeat (SSR).

16. An inert support comprising a plurality of allele specific oligonucleotide probes (ASOs) bound to the surface of the support as claimed in claim 15, in which the inert support is nitrocellulose, polyethylene, polypropylene, silicon or glass.

10

17. A kit for the determination of an organisms genotype comprising an inert support as defined in claim 15 or claim 16 and a means for detection the hybridisation of an ASO to a target DNA sequence.



1/5

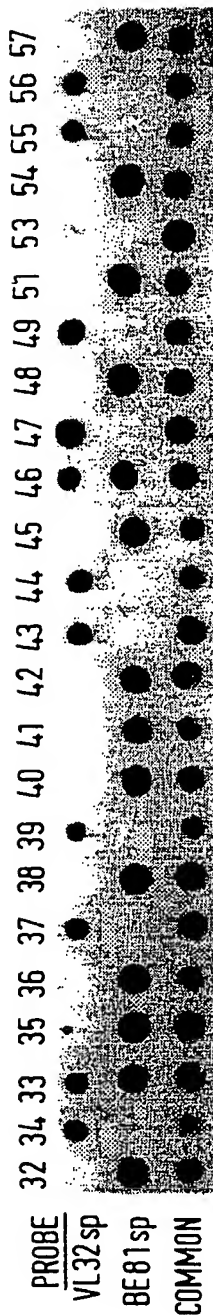
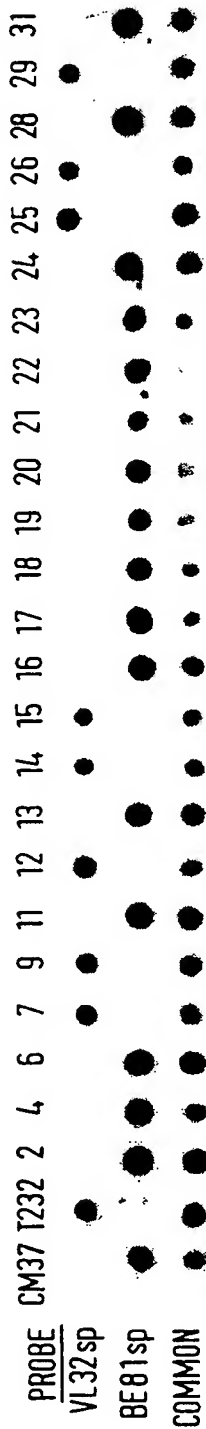


FIG. 1

2/5

1. Maize

MICROSATELLITE MAC01F07: FORWARD PRIMER

M017F ACATGTGAGACGTCCCTCCCGGTAGTGACAGTACTACCGG-----AATCCGTAGCAGCGGGGGGGCT---CAATGTTTCTGCAGCA  
 D1143F ACATGTGAGACGTCCC--TCTCCGGTAGTGACAGTACTACCGG-----AATCCGTAGCAGCGGGGGGGCAATGTTTCTGCAGCA  
 C0125F ACATGTGAGACGTCCCTCCCGGTAGTGACAGTACTACCGG-----AATCCGTAGCAGCGGGGGGGG--CAATGTTTCTGCAGCA  
 F7F ACATGTGAGACGTCCC--CTCCGGTAGTGACAGTACTACCGG-----AATCCGTAGCAGCGGGGGGGG--CAATGTTTCTGCAGCA  
 CM37F ACATGTGAGACCTCCCCCTCCGGTAGTAACA TTA TACTACCGA-----AATCCGTAGCAGCGGGGGGGGTG--CAATGTTTCTGCAGCA  
 C0159F ACATGTGAGACGTCCC--CCTCCGGTAGTGACAGTACTACCGGGCGG-AATCCGTAGCAGCA  
 F2F ACATGTGAGACGTCCCCCTCCGGTAGTGACAGTACTACCGGGCGG-AATCCGTAGCAGCA  
 T303F ACATGTGAGACGTCCCCCTCCGGTAGTGACAGTACTACCGGGCGG-AATCCGTAGCAGCA  
 B73F ACATGTGAGACGTCCCCCTCCGGTAGTGACAGTACTACCGGGCGG-AATCCGTAGCAGCA  
 B37F ACATGTGAGACGTCCCCCTCCGGTAGTGACAGTACTACCGGGCGG-AATCCGTAGCAGCA  
 B14F ACATGTGAGACGTCCCCCTCCGGTAGTGACAGTACTACCGGGCGG-AATCCGTAGCAGCA

Single  
Large

FIG. 2

3/5

## SEPTORIA SEQUENCES 312F

RA11F-CTCTGGATGCGCGNTAACGTCCGGCAACAGCAGTAGAGCGTTCGCAGCGTTGAACGCTTGAAAGAGCGTGATATCCGCGTCGTCCATGGGGGTGTGTATCGAACGAG  
RA3F-CTCTGGATGCGCGNTAACGTCCGGCAACAGCAGTAGAGCGTTCGCAGCGTTGAACGCTTGAAAGAGCGTGATATCCGCGTCGTCCATGGGGGTGTGTATCGAACGAA  
RA5F-CTCTGGATGCGCGNTAACGTCCGGCAACAGCAGTAGAGCGTTCGCAGCGTTGAACGCTTGAAAGAGCGTGATATCCGCGTCGTCCATAGGGCGTGTGTATCGAACGAA  
RA6F-CTCTGGATGCGCGNTAACGTCCGGCAACAGCAGTAGAGCGTTCGCAGCGTTGAACGCTTGAAAGAGCGTGATATCCGCGTCGTCTATAGGGGTGTGTATCGAACGAA

RA11F-GGCGGCGGTGGTGGCGGCGGCGG-----TGAGTATGAGATGTGAACGT  
RA3F-GGCGGCGGTGGTGGCGGCGGCGG-----TGAGTATGAGATGTGAACGT  
RA5F-GGCGGCGGTGGTGGCGGCGGCGGCGG-----GCA-----GCGGCGGTGAGTATGAGATGTGAACGT  
RA6F-GGCGGCGGTAGTGGCGGCGGCGGTAAAGTATAGATGTAAACGCGGCGGTGAGTATGAGATGTGAACGT  
\*\*\*\*\*

FIG. 3

## PIC SSR Sequence Alignment

Pig SH524.1: CTCCTGGAGCTAGATTTT-----GTTGTGTGTGTGTGTGTCTTTTGGCCATTTCCTTGGGCCGCTCTTC  
Pig SH525.1: CTCCTGGAGCTAGATTTTGTGTGTGTGTGTGTGTCTTTTGGCCATTTCCTTGGGCCGCTCTTC

AGCATATGGAGGTTCCCAAGGCTAGGGGTCCAATTGGAGCTGTAGCGCTGGCCCTGCACCAAGCCACAGCAACCGCGGCTCT  
AGCATATGGAGGTTCCCAAGGCTAGGGGTCCAATTGGAGCTGTAGGCACTGGCCCTGCACCAAGCCACAGCAACCGCGGCTCT

GAGCTTGTTGTC TCGGACCTACACCAACTC  
GAGCCGTGTTGTC TCGGACCTACACCAACTC

FIG. 4

NUMBER	NAME	SEQUENCE
1.	M017.1	5'ACGTCCCCCCTCCGG
2.	M017.2	5'TACGGGAATCCGTAG
3.	OH43.1	5'ACGTCCCCTCCTCCGGT
4.	OH43.2	5'GGGTGGGGGGGCAA
5.	C0125.1	5'GGCGGGGGGGGCAATG
6.	C0159.1	5'CGCGCGGAATCCGTAG
7.	CONTROL	5'TCGTTCGGTCCCATGAA

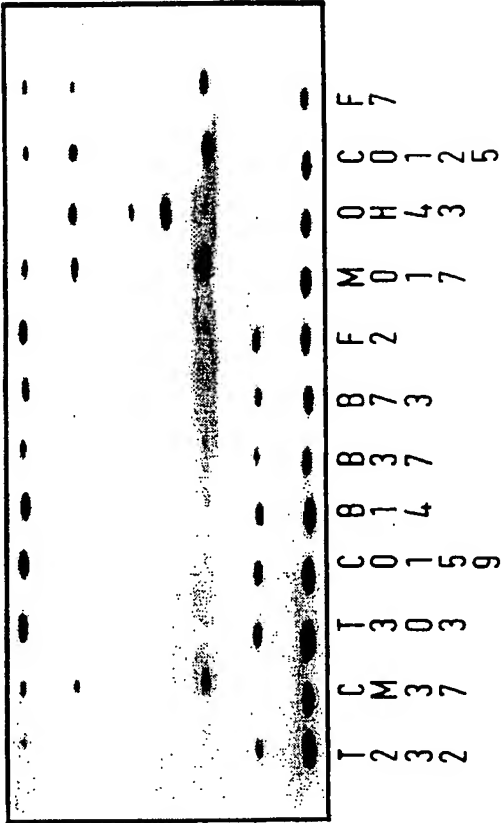


FIG. 5

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 98/01940

A. CLASSIFICATION OF SUBJECT MATTER  
IPC 6 C12Q1/68

According to International Patent Classification(IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	GRIMALDI AND CROUAU-ROY: "MICROSATELLITE ALLELIC HOMOPOLYMER DUE TO VARIABLE FLANKING SEQUENCES" J. MOL. EVOL., vol. 44, 25 February 1997, pages 336-340, XP002082097 see the whole document ---	1-17
A	US 5 633 134 A (SHUBER TONY) 27 May 1997 see the whole document ---	1-17
A	PROSSER J: "DETECTING SINGLE-BASE MUTATIONS" TRENDS IN BIOTECHNOLOGY, vol. 11, June 1993, pages 238-246, XP000601507 cited in the application see the whole document ---	1-17
-/--		

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

## \* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

26 October 1998

Date of mailing of the international search report

11/11/1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Hagenmaier, S

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 98/01940

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category <sup>a</sup>	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 95 15400 A (UNIV JOHNS HOPKINS) 8 June 1995 see the whole document ----	1-17
A	DE 195 25 284 A (INST PFLANZENGENETIK UND KULTU) 2 January 1997 see the whole document ----	1-17
A	CA 2 013 430 A (LITT,M.) 29 September 1991 see the whole document ----	1-17
A	ROTHUIZEN AND RAAK: "RAPID PCR-BASED CHARACTERIZATION OF SEQUENCES FLANKING MICROSATELLITES IN LARGE-INSERT LIBRARIES" NUCLEIC ACIDS RESEARCH, vol. 22, no. 24, 1994, pages 5512-5513, XP002082098 see the whole document ----	1-17
A	LIPSHUTZ R J ET AL: "USING OLIGONUCLEOTIDE PROBE ARRAYS TO ACCESS GENETIC DIVERSITY" BIOTECHNIQUES, vol. 19, no. 3, 1 September 1995, pages 442-447, XP000541924 see the whole document -----	15-17

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/GB 98/01940

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
US 5633134	A	27-05-1997	NONE	
WO 9515400	A	08-06-1995	NONE	
DE 19525284	A	02-01-1997	WO 9701567 A EP 0835324 A	16-01-1997 15-04-1998
CA 2013430	A	29-09-1991	NONE	